Our responses to reviewer #2

The observations and comments by the reviewer are greatly appreciated. We find this review to be highly objective and constructive. This will go a long way in improving the quality of our submission. We have generally agreed with most of the observations made by the reviewer in the manuscript and the additional comments. The reviewers comments in both comments are generally similar and here attached are our responses (in blue) to the issues raised (in red) by the reviewer:

Title and abstract

We appreciate this feedback. We certainly agree that emphasis on phenology was not adequate. We will address the issues raised.

Methodology for evaluating RS evaporation products

Our initial reason for use of the FLEX-Topo and TerraClimate as reference products at monthly scale was to help understand the dry season evaporation using water balance approach since this is generally not possible with the general water balance approach (i.e., P-Q). However, the observations made by the reviewer are valid. We believe this part of the study can be adjusted. The FLEX-Topo and TerraClimate will not be used as reference data but be treated like the rest of the evaporation products. The FLEX-Topo and TerraClimate will also be compared to the general water balance just like the rest of the products.

E_{wb} was calculated for calendar years, not hydrological years which may influence the over year storage variations, better to reduce this additional source of error by comparing the hydrological year

This is a valid observation. We initially considered this approach and actually compared the results for the calendar and hydrological years. There were no significant differences in the results. Our decision to settle for the calendar year was to help with the understanding of the phenology. The hydrological year for the Luangwa Basin is from September to August. The phenological processes, of for example, leaf fall and leaf flush (and other accompanying processes) as highlighted by Frost (1996), Fuller (1996) and Chidumayo 2001, can start early and end early, start late and end late going up to November/December depending on the rainfall in the preceding rain season. We wanted to observe the pattern across phenophases and not necessarily hydrological years. In this instance, our view was that phenology was better observed using calendar and not hydrological years.

The selection of precipitation product used for the Ewb calculation seems to be based on 8 observations in three locations (Figure A1), this is a very limited validation and the data should therefore be considered with certain uncertainty

and not seen as an absolute reference to which the other products need to adhere to (you already indicate that the absence of the over-year storage is reducing the value of Ewb as a validation product.

This is an important observation. Our initial plan was to use the average of an ensemble of five precipitation products. However, we felt this would simply mean aggregation of error. This is why we decided to use one product that we deemed, after very limited validation, would be representative of basin precipitation. However, as validly observed by the reviewer, this approach has short-comings. We feel the use of an ensemble of satellite precipitation product(s) will reduce the error and possibly give better results as has been recommended by Asadullah *et al.* (2008). This can easily be done.

The explanation of the Ewb validation data (incl selection of the precipitation product) can be presented fully in the methodology (and not have two sentences at the start of the results sections, which seem a bit out of place/ duplication “sensitivity of the precipitation product”).

This is a valid observation. This will definitely improve the quality of information and presentation for this section.

To me it is not clear what the purpose of the trend analyses and correlation analyses is, the reference data set (Ewb) has more variation than the evaporation products being evaluated, what does this mean? The fact that there is no significant trends (for 12 years of data, which is a too small data set) for the different products, what does this mean?

We compared the products at basin scale across a 12 year time space. Existence of a trend would mean “significant” differences in hydrological conditions during the period. Comparison is supposed to be under “similar” (no trend).hydrological conditions. The general ideal behind this approach is that existence of trend (-/+) would have necessitated assessment of the change point (s) so that comparison is done in segments with no trend(s).

Trend analyses at monthly timescales seems to be done for the entire dataset (144 months) instead of comparing similar data (Jan alone), if you do take the full time series, the trend analyses is influenced by the seasonality of the data, how do you account for that?

We accounted for seasonality by performing seasonal adjustments on the monthly time series in python. However, like the reviewer has suggested, this can be done based on individual months or phenophases should there be need.

TopoFlex and TMC seem to be used as standard to compare the other products against (eg figure 6 & 10) however later in the paper it becomes clear that these products also have their limitation to model evaporation. For example figure 7 shows the mismatch between the NDVI and the spatial patterns of the TopoFlex and TMC models.

This is an extremely important observation and has been discussed above. Please refer to our response above. This part needs to be revised.

The higher variability observed by SSEBop and WaPOR shows the basins is most likely due to the higher resolution of the data, including being able to identify water bodies with high evaporation.

We looked at this and thought the same. We considered resampling the products. However, we used the native resolutions as they are made available (read to use) on the various platforms. Your observation is correct and highlights the ability of higher resolution products to capture evaporation dynamics due to capacity to differentiate various land cover features. However, the other question that should be asked is why MOD16 with a higher resolution than SSEBop and FLEX-Topo, appeared to underestimate evaporation in some periods. Also MOD16 with a higher resolution than WaPOR appear to have higher evaporation values between March and June in the dense forest land cover. This interchange in behaviour, to a large extent, made us think that the spatial resolution did not have a significant role.

Not clear to me how the NDVI spatial pattern is used to evaluate the spatial patterns of the different evaporation products? The text now seems to be descriptive.

NDVI is used as proxy for availability of green vegetation. This component was aimed at observing spatial relationship between evaporation and NDVI. As correctly observed by the reviewer a more detailed comparison could be useful than the descriptive type used earlier.

The first analyses done are to compare the entire basin evaporation data against the Miombo phenophases, however we only find out in the discussion that only 60% of the basin is covered by Miombo woodlands (at least if we assume open forest classification is all Miombo woodlands). What is the justification for this assumption and would you assume that the entire basin would respond in a similar way as the Miombo woodlands?

The Luangwa Basin, based on the biomes map of the Miombo region, is about 70 percent Miombo Forest (dry and wet). The other 30 percent in largely in low lying areas that are mainly grass and shrub land. We observed major differences in model performance in the dry season. Our assumption was that the grasslands/shrub-lands normally dry out during the dry season with minimal evaporation. Using the evaporation maps we observed this assumption to be generally true. We assumed much of the dry season evaporation would come from the Miombo Forest. What is true, as the reviewer has observed, is that the moisture feedbacks of non Miombo areas is definitely not the same as that of the Miombo Woodland. Non-Miombo areas and water bodies from all products can easily be masked out and then make the comparisons with only Miombo Forest. Doing this will improve the results of the comparisons of the various evaporation products.

Comparing the five selected locations (Figure 8)

The six selected locations vary in size as can be seen from the number of pixels used from the WaPOR data (256-2304), however for the low resolution data, each time only one pixel is considered in the analyses. It can be assumed that for the smaller areas, these pixels overlap with the surrounding areas, which could have different land cover types which may have influenced the temporal signature. How has this been taken care of? In table A5 you indicate the same number of observations for each of the locations, how did you aggregate the WaPOR data (average?) to compare it as one value against the other datasets?

Comparison between coarse and fine resolution products is always a challenge especially when dealing with specific land cover classification. An objective comparison would be to have the same resolution for all products. But this is difficult to achieve. This is because, resampling the high resolution products to match the coarse resolution products or vice versa creates a new set of undesirable and unquantifiable errors. The pixel overlap between features is unavoidable when coarser resolution products are used. Even fine resolution products like WaPOR cannot completely avoid feature overlaps because of the characteristics and heterogeneity in the land cover even at smaller scales. For instance the individual tree canopies and entire forest canopy do not have 100 percent closure. What is termed as forest land is not completely forest only and the spectral signature includes other features such as bare soil and plant litter. WaPOR's 250 m by 250 m resolution includes different vegetation types (i.e., trees and grass) and bare land/soil. In this study we used the products as is since we did not run the models. We used the native resolutions in order to first of all not introduce errors by resampling the resolutions of the products. Secondly we wanted to observe if there would be significant differences in the results using the native resolutions. These products are normally used in their native resolutions as provided on source platforms. This is why we felt comparison should also be using the native resolutions so that the comparison outcome would show whether a higher resolution product generally give better results than coarser resolution product.

We created polygons/shape-files for each land cover type. Using the shapefile(s) we extracted the evaporation values within that area. Normally, for MOD16, SSEBop and WaPOR, even TerraClimate had all pixels falling within the shapefile. Overlap with other pixels was within the same land cover types. We simply aggregated the evaporation values for each pixel, at native resolution, for each evaporation product. The challenge was with the FLEX-Topo and GLEAM especially with the water body and crop/agricultural land classifications which normally were smaller.

Discussion

In the discussion section a lot of new data is presented, this is not normally a good place to present new data and analyses. For example figure 11 presents new data, but to me it is not very clear how this contributes to understanding how well the different evaporation data products are able to monitor the Miombo woodlands. Similarly the sections with the explanations on how SSEBop and WaPOR perform contain a lot of new information on how the Miombo woodlands work and which is used to confirm that the evaporation observed by SSEBop and WaPOR at the end of the dry season are not unrealistic. In my opinion it would have been helpful if this information was presented upfront, including figure 13 with the land cover classes observed in the Luangwa basin.

This observation is greatly appreciated. We have taken note.

Specific comments:

You categorise the remote sensing evaporation products into energy balance models (EBM), however the WaPOR methodology is not a surface energy balance model, instead it uses Penman-Monteith (ETLook) for estimating evaporation.

This a valid observation. It will be looked into.

Update graphs to remove the digits (eg 100.0 should be 100)
Figure 5 title of 5B monthly average (year 2009-2020) and not 2019-2020

This is noted

We hope the responses we have provided have largely addressed the issues raised by the reviewer. We are very much willing to engage further and improve the quality of our submission.

References

Asadullah, Anita, Neil McIntyre, and Max Kigobe. 2008. "Evaluation of Five Satellite Products for Estimation of Rainfall over Uganda." *Hydrological Sciences Journal* 53 (6): 1137–50. https://doi.org/10.1623/hysj.53.6.1137.

Chidumayo, E. 2001. "Climate and Phenology of Savanna Vegetation in Southern Africa." *Journal of Vegetation Science* 12 (3): 347. https://doi.org/10.2307/3236848.

Frost, P. 1996. *The Ecology of Miombo Woodlands*. *The Miombo in Transition: Woodlands and Welfare in Africa*. http://books.google.com/books?hl=nl&lr=&id=rpildJJVdU4C&pgis=1.

Fuller, Douglas O. 1999. "Canopy Phenology of Some Mopane and Miombo Woodlands in Eastern Zambia." *Global Ecology and Biogeography* 8 (3–4): 199–209. https://doi.org/10.1046/j.1365-2699.1999.00130.x.